

November 2006

Moving Towards Data-Driven Development: WorldCat Collection Analysis

Alice Sneary

OCLC, snearya@oclc.org

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

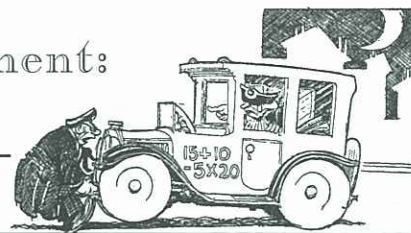
Sneary, Alice (2006) "Moving Towards Data-Driven Development: WorldCat Collection Analysis," *Against the Grain*: Vol. 18: Iss. 5, Article 12.

DOI: <https://doi.org/10.7771/2380-176X.5012>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Moving Towards Data-Driven Development: WorldCat Collection Analysis

by Alice Sneary (OCLC Creative Services Analyst) <snearya@oclc.org>
(or It's All Good blogger at scanblog.blogspot.com)



The comedian **Steven Wright** once said, “You can’t have everything... where would you put it?” And sometimes it seems like that joke holds true with data about our library collections, too. It should be easy to see current information at any time, on every item in the collection — or in someone else’s collection. But if you had all that data and metadata...if you had “everything”...where would you put it? And how could you keep it reasonably current?

Until recently, that’s been the conundrum. Traditional collection data and analysis tasks have been criticized for being too informal and qualitative. But the rigorous methodology and volume of data required for truly quantified, authoritative analysis have been both prohibitively expensive and time-consuming. Management and administration may want to analyze “everything,” but real-life staffing and budget restraints stood in the way.

Until now. Since its launch in March 2005, more than 500 libraries have discovered that they have somewhere “to put everything” in their collections so they can analyze it, compare it and plan for the future. All their collection data and metadata now resides in **WorldCat** as part of the **WorldCat Collection Analysis** service from **OCLC Online Computer Library Center**.

Participating libraries know what’s there, and what’s NOT there, for their collections. They know what titles and subjects are good candidates to acquire, and what should be weeded out. They can compare themselves against peer institutions and authoritative lists, using the more than 1.3 billion items in **WorldCat** in a whole new way.

What is WorldCat Collection Analysis?

WorldCat Collection Analysis is a Web-based service that allows libraries to analyze all the materials they have in their collections that are cataloged in **WorldCat** — their “holdings.” Libraries can evaluate their holdings in multiple ways — by format, subject, language, audience or publication date, to name a few. Then they can generate an easy-to-read (and visually intuitive) report as a table or a graph, ready to drop into a printed report or electronic newsletter. Library staff access **WorldCat Collection Analysis** through a tab on the **OCLC FirstSearch** service, which means it’s a familiar, comfortable interface for many librarians to navigate.

Any library with holdings in **WorldCat** can use **WorldCat Collection Analysis**: community college libraries, academic libraries, public libraries, special libraries such as law, art, medical and theological — even government libraries use the service. A sizable number of subscribers (23 percent) have less than 100,000

holdings. About 17 percent of subscribers have more than one million. In other words, the service has value for any library, of any size.

A library can compare its collection against other collections, too. Once the individual library analysis is complete, a library can do a peer-to-peer comparison against up to five other libraries’ collections at a time — regardless of whether the other libraries use the service.

Predefined peer comparisons are also available, such as the **Google Book Search Library Partners** (the “**Google Seven**”), the top 20 **ARLs**, the top 10 law schools and the top executive MBA programs — as determined by external sources such as *U.S. News & World Report*. Or, you can compare against known authoritative source lists. Some examples include *Outstanding Academic Titles*, *Booklist*, *Books for College Libraries*, 3rd edition and more.

Visualization tools such as charts and graphs help library staff see patterns and draw insights from **WorldCat Collection Analysis** data. A library can see itself as a whole or how it stands in relationship to other libraries, at a glance. Data visualization enables a “drill-down” effect to see successively more granular views.

Why Data Matters

WorldCat Collection Analysis is important because it provides a powerful insight into a library’s collection strengths — and its weaknesses. These insights, in turn, can drive accreditation discussions and funding presentations, dovetail with census data predictions and more.

To academic libraries

An analysis of your collection could help align day-to-day collection development goals with the strategic goals of the university. It could be as simple as continuing accreditation, but it could be much more than that, too.

For example, one goal might be to become the leading university for graduate-level bioengineering studies in the eastern United States. The library can do a comprehensive analysis on the bioengineering materials available, and then compare its holdings with collections in other leading institutions. **WorldCat Collection Analysis** provides the tools to make and communicate smart acquisition decisions and develop funding requests to support the university’s long-term goals.

There has been an increased emphasis in many academic institutions for faculty to do more original research, speak publicly and publish. **WorldCat Collection Analysis** can help collection development staff develop library collections that support faculty efforts to reach these goals.

To public and special libraries

Public libraries also benefit from learning more about what’s available and accessible to users with **WorldCat Collection Analysis**. Publics have also used it to determine multilingual strengths, subject gaps and where they might find some additional shelf space with a judicious weeding of titles. Special libraries can use it to determine age range of specific collections, or to target potential materials for digitization.

To consortia — geographic or type-specific

Consortia have started to use **WorldCat Collection Analysis**, too. They can see the materials in their geographical area — or can look at similar special-interest collections (**Braille**, for example) and move to complement each other’s holdings, rather than duplicate them.

Reciprocal borrowing arrangements can be made and each consortial member library saves money on duplicative materials and continues to provide users with quick access to physical materials beyond local walls. We’ll look at some surprising results of consortia analysis, the **TRLN Collection Analysis Project**, later on.

An Independent Source of Data

WorldCat Collection Analysis has emerged as such a valuable tool because it gives libraries an objective look at the facts behind their collection development decisions over the years. Back to the beginning, in fact.

Libraries subscribe to the service on an annual renewal basis, which comes with unlimited report usage. That means a library can take a baseline snapshot of its collection, fill gaps, weed, strengthen subject areas and run a report each month to check against goals. It’s a great way to benchmark and then measure progress.

Erlene Dudley, Director of **William Woods University Library** in Fulton, Missouri, explained the importance of the service in helping to reach their goals when she says “We’ve been able to establish benchmarks to build on...and [**WorldCat Collection Analysis**] has been extremely useful for developing budgets because it’s an independent, objective source of data.”

With the rising cost of acquisitions, library boards across the country demand more insight into library budgets, and a shift towards data-driven models of collection development has been seen nationwide. With **WorldCat Collection Analysis**, library staff members can show where their library is unique and where the opportunities lie for improvement. Boards can base decisions on facts and comparisons to similar libraries, rather than guesses or intuition.

As one satisfied librarian told us in an anonymous survey, *continued on page 56*

mous user survey, "Being able to compare our collection with peer libraries is amazing. I have only begun to imagine the possible uses for the information."

The WorldCat in WorldCat Collection Analysis

When a library uses the service, it puts the library's materials within the larger context of the 71 million items (records) in **WorldCat**. New holdings pour in every day, as the database continues to expand beyond the current 1.3 billion mark. In fact, an **OCLC** member library adds a record to **WorldCat** every 10 seconds, on average.

The constant rate of expansion makes batchloading a critical component to a library's successful use of the service. The more data your library puts into **WorldCat** — the more data your library will get out of **WorldCat**, in terms of collection analyses and more. Libraries that subscribe receive a complimentary retrospective batchload of holdings into **WorldCat**.

Batchloading gets metadata into **WorldCat**, so your library can analyze the results. But your users are more likely to be interested in what materials they can get out of **WorldCat**. That is, **WorldCat Collection Analysis** can also help optimize your library's collection for **WorldCat.org**.

Staff may look at subject area gaps and realize that eBooks or eJournals would be the best way to connect users immediately with the content they're looking for. Run the NetLibrary and eBooks comparison list to help your library supplement print materials with electronic.

Through **WorldCat Collection Analysis**, a librarian can see at a glance what eContent, digital materials and print citations the library makes available through **WorldCat.org** on the open Web — and how it stacks up against peer libraries.

Success Stories

So who are the people using this service? **Glenda Lammers**, Global Product Manager of **WorldCat Collection Analysis**, provides some insight into these 500 "data-driven libraries."

"I am always amazed at the ingenuity of our libraries. They are always pushing the service and finding new ways to mine the data from **WorldCat**. Since the 2005 release, there have been 1,643 analyses done on the service, with peer comparisons being by far the lion's share at 73 percent of total analyses run."

That means libraries have used **WorldCat Collection Analysis** multiple times in the past 12 months. Historically, libraries have been more interested in knowing how they stack up against each other, although the interest in measuring against an "ideal" authoritative source has started to take off.

Lammers confirms this, as she tells me that Authoritative List comparisons were only added to the service in June 2006, and as of this writing (early September), there have already been 68 reports run in about as many days. The de-

against the grain people profile

Alice Sneary

Creative Services Analyst
Creative and Branding Services, OCLC
6565 Frantz Road, Dublin, OH 43017
Phone: (800) 848-5878 x 6016 or (614) 764-6016
Fax: (614) 718-7235 <snearya@oclc.org>

BORN & LIVED: Born in Okemah, OK. Now live in Dover, New Hampshire. Have previously lived in Alva, OK; Sewanee, TN; London, England; Exeter, England; Honolulu, HI and Columbus, OH.

EARLY LIFE: I grew up on a farm in northwest Oklahoma, where my dad was an Episcopal priest/farmer and my mom was a music teacher. I did gymnastics and sang in the choir. I was lucky that my extended family all lived very close to each other, growing up.

FAMILY: My husband is a graduate student, working on his Ph.D. at **UNH** in Durham, NH in ocean engineering.

EDUCATION: MA in Literary Criticism and Theory from the **University of Exeter** in Exeter, England. BA in English Language and Literature from the **University of the South** in Sewanee, TN.

FIRST JOB: I was a researcher for a legal defense firm in Hawaii on the largest civil case in the State of Hawaii history. I added metadata/cataloged discovery evidence into a database.

PROFESSIONAL CAREER AND ACTIVITIES: I've concentrated on library marketing and advertising for **OCLC** for the past six years. I started the *It's All Good* blog with **Alane Wilson** and **George Needham**. *IAG* bloggers now also include **Chrystie Hill** and **Eric Childress**. I've also been lucky enough to do creative work on **OCLC's** library advocacy advertising projects.

IN MY SPARE TIME I LIKE TO: Tell wild stories. Laugh loudly. Play soccer. Go to movies. I am a recovering **Frank Lloyd Wright**/modern architecture addict.

FAVORITE BOOKS: For business, I really like *The Art of Innovation* and *The Ten Faces of Innovation* by **Tom Kelley**.

PET PEEVES/WHAT MAKES ME MAD: Being told that something is a great idea, but that it's too risky for an audience.

PHILOSOPHY: Be Bold! Plus, I like to take an anthropological approach. Watch where people go, see what they do and then build your services/products around that. Don't try to change behaviors — offer out a more convenient alternative with a good experience and people will fit you into their lives — rather than asking them to fit around you.

MOST MEANINGFUL CAREER ACHIEVEMENT: Surely it hasn't happened yet?

GOAL I HOPE TO ACHIEVE FIVE YEARS FROM NOW: I'll have written a book, done some wildly creative work and put libraries back in mainstream public consciousness as a desired destination both online and in print.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: In five years, I think we'll see much more social stuff going on with library technology. I think we'll also see more instances of library partnerships — mini-library outposts or kiosks at airports, hospitals, doctors offices, hardware stores, grocery stores, coffee shops.

Wherever people need information — anything from "what does it mean that my child was just diagnosed with autism" to "how do I build a deck on my house" — the library has resources that can help. The key is making sure that the library is presented as an additional trusted information source, where and when the person needs the information. 🐾



mand for comparisons and comparison lists only continues to grow.

William Woods University

The private, liberal arts institution in Fulton, Missouri, started to use **WorldCat Collection Analysis** and went from peer comparison to

purchase in little over three months. As Library Director **Erlene Dudley** explains, "Simply saying we have this many titles in History doesn't do much. We also wanted to see how we compared to other institutions to identify what we needed."

continued on page 58

Dudley was looking for context, and she got it. "We realized that our NetLibrary eBooks filled the subject gaps we identified. We then found people were using them frequently. This opens up our acquisition thinking."

William Woods used the service to do peer-to-peer comparisons and bolster academic subject strengths.

Southeastern Oklahoma State University

The state school in Durant, Oklahoma, supports more than 4,000 students in 19 undergraduate programs and five graduate programs. But the library struggled to maintain a consistent collection development scheme. "We never knew with certainty how well our titles supported the academic program," says Collection Development Librarian Susan Webb.

With **WorldCat Collection Analysis**, they had the necessary data to make curriculum-supporting decisions to assist an accreditation review. **WorldCat Collection Analysis** opened her eyes. "You don't realize how heavily a collection might be slanted toward a department until you look at it in detail. And you can't dispute it because it's there in black and white," says Webb.

For **Southeastern**, using the individual collection analysis portion of the service contributed to the institutional accreditation procedures.

Virginia Tech

With more than 2.2 million titles in its libraries, **Virginia Tech** had lots of questions about its collection — especially because it was spread among a main library and three specialty branches, explained Paul Metz, Director of Collection Management. **Virginia Tech** is a member of **ASERL**, the **Association of Southeastern Research Libraries** — the largest regional library consortium in the country. Through **ASERL**, **Virginia Tech** could analyze its own collections, as well as compare them to 13 doctoral institutions that were custom-selected as a peer group for **Virginia Tech**.

With the new service, Metz reports he is able to "speak empirically, comparatively and in detail, whereas without it I could only have been speculative or very general." He has led the effort to look at collections on a large scale, down to a very granular level. "It gives us the ability to make smarter decisions about what's needed in our collections."

Virginia Tech used **WorldCat Collection Analysis** to gain an overall sense of what's in its collections and determine gaps and overlaps with peer institutions.

Triangle Research Libraries Network

The four universities that make up the **Triangle Research Libraries Network (TRLN)** include **North Carolina State University**, **Duke University**, **North Carolina Central University** and **University of North Caro-**

against the grain people profile

Glenda Lammers

Global Product Manager
Cooperative Collection Management, OCLC
6565 Frantz Road, Dublin, OH 43017
Phone: 800-848-5878 x.6062 or (614) 764-6062
<glenda_lammers@oclc.org>

BORN & LIVED: Born in Mobile, AL; live in Columbus, OH. Lived in Birmingham, Decatur and Carrollton, AL; Newfoundland, Canada; Ashville, Ohio; Dallas and Abilene, TX; Yuma and Tucson, Arizona; Centreville, VA and Olympia, WA.

EARLY LIFE: My father was in the Air Force so some of the places listed above I did not choose! Played basketball everyday after school on dirt court behind my house.

EDUCATION: MSLIS from **University of North Texas**, 1997; BA in Accounting from **Birmingham-Southern College**, 1981.

FIRST JOB: Doing inventory at the local hardware store in rural Alabama.

PROFESSIONAL CAREER: My library career includes: Product Manager for collection analysis services at **OCLC** since 2000; Library Reference trainer at **Amigos Library Services**; Reference and Collection Development Librarian at **Mesquite Public Library**. My accounting career includes: Auditor for the **Pension Benefit Guarantee Corporation**; Auditor for the **US Army**.

IN MY SPARE TIME I LIKE TO: Run in 5K's, 10K's and half-marathons. I'm training for my first (and possibly last) marathon. Hopefully by the time you read this, I'll have completed it!

FAVORITE BOOKS: Anything by Jon Krakauer.

PET PEEVES/WHAT MAKES ME MAD: People saying "that's not my job" and/or saying "that won't work" before even trying.

PHILOSOPHY: Just because you Can build it doesn't mean you Should. Always consider the user.

MOST MEANINGFUL CAREER ACHIEVEMENT: Contributing in the development of a tool useful for libraries in their day-to-day collection development activities.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: Cooperation will be even more important to determine what core materials in print and electronic materials are added to collections, what is maintained in off-site storage facilities and what libraries discard from their collections.



lina. A task group was formed to oversee a consortial project using **WorldCat Collection Analysis** to learn about the overlap, uniqueness and gaps in the collections of **TRLN** libraries, especially when viewed as one aggregate collection.

The initial findings of the study were quite surprising. Based on the 5,158,309 **OCLC** holdings included in the overlap study, only 2 percent of the records overlap at all four institutions. Seventy-one percent of the aggregate **TRLN** collection, it turns out, are unique to a single institution. When viewed as a whole, then, the collection in this geographic area is quite strong indeed.

Cooperative, collaborative collection development was what made **WorldCat Collection Analysis** so appealing to the **TRLN** consortia.

Surprising Uses: Preservation, Digitization, Music and Languages

Looking at a library's collection for gaps, overlaps, strengths and weaknesses is one of the main uses for **WorldCat Collection Analysis**. But libraries have also used it in surprising ways. For example, Shirley Baker, Vice Chancellor, Information Technology and Dean, University Libraries of **Washington University** in St. Louis, Missouri, is using it to drive a preservation and digitization project.

Her library staff has compared their collection with two of the libraries that participate in the **Google** book digitization project. They found that 150,000 public domain works held by their library would eventually be freely available through **Google**, but that

continued on page 60

Moving Towards Data-Driven from page 58

400,000 works held by **Washington University** are NOT being digitized in the **Google** program. Now they can use **WorldCat Collection Analysis** to see which of the 400,000 titles are uniquely held in their collection. These titles — **Baker** estimates to be somewhere between 7,000 and 30,000 — will form the basis of their preservation and digitization efforts.

Figures 1 and 2 show how a library might make more use of **WorldCat** data in unexpected ways to drive decisions.

For example, **figure 1** shows how **WorldCat Collection Analysis** can generate a report on the distribution of languages for materials published from 1600-1699.

Should the yearly medievalist colloquium be held at this library, the Latin scholars could rejoice... and the library would know to request some additional Dutch, German or Italian materials before the event. Easy-to-read graphs show a collection snapshot, and staff can "drill-down" for ever more granular detail.

Figure 2 selects by format. The analysis looked for sheet music, musical scores and electronic music. Again, the graph demonstrates a clear pattern of build-up in the 1970s and then a marked drop-off. Why is the music program not growing on campus? Perhaps a funding decision in the 1980s had something to do with it.

What's next for WorldCat Collection Analysis?

Glenda Lammers has big plans for the next release of **WorldCat Collection Analysis**, due in late 2006.

"We're working on incorporating inter-library loan data from **WorldCat Resource Sharing** into **WorldCat Collection Analysis**. It will help libraries see where their collections are augmented through resource sharing, and whether or not the interest is temporary or sustained."

The enhancement, available to current subscribers at no charge, will also show if libraries are lending out materials often enough to justify space on the shelves. It will show what materials would be good candidates for off-site storage, as well as identify where a strong collection may be aging. That is, a library may have a strong collection in nursing, but the majority of the materials may be more than five years old. Data from **WorldCat** that includes ILL requests will provide an even more comprehensive approach to analyzing and comparing a library's collection. **ILLiad** and **Fretwell Downing VDX** statistics will be included in a



Figure 1: A graph of items held by the library in Language, Linguistics and Literature, published between 1600-1699, sorted by language.

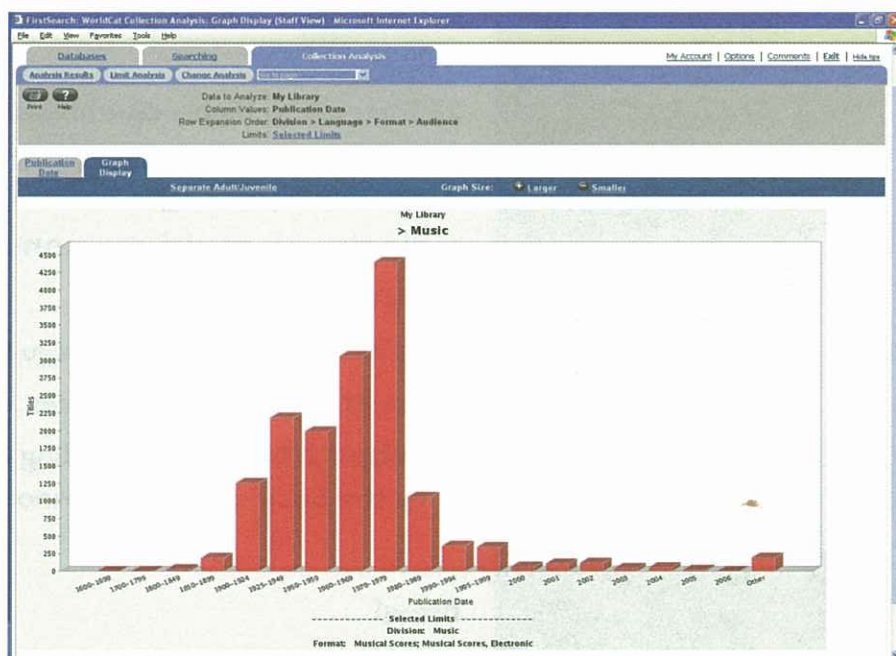



Figure 2: A report of holdings by format. This graph shows musical scores broken out by decade.

future enhancement, due in spring 2007.

"You'll be able to see the patterns in the ILL requests," explains **Lammers**. "It will give libraries an even clearer picture about what's in their collections and

how it compares with what users are looking for."

And getting users what they want is, in the end, the whole point of collecting all this data and analyzing it more efficiently. 

Rumors from page 52

helpful in locating hotel rooms for our attendees. Thank you, **Michael!**

Speaking of helpers with the **Charleston Conference**, many, many, many people have helped us. First and foremost, on the local front, are **Regina Semko** <semkor@cofc.edu> who once worked at the **Lightsey Conference Center** and knows all about running **Conferences** and **Leah**

Hinds <leah_hinds@hotmail.com>, who used to work with **Regina** and is learning the ropes. Our other indispensable on-site helper is **David Lyle** who is a fabulous techie and can solve any problem in the world. **David** will also do any task we

continued on page 78